KLUCZOWE POJĘCIA DOTYCZĄCE DATA SETU

**✅ Medicare**

* **Co to:** Rządowy system ubezpieczeń zdrowotnych w USA, głównie dla osób powyżej 65 roku życia.
* **W datasetcie:** To źródło finansowania usług medycznych, które są rozliczane w roszczeniach (claims).

**✅ Provider**

* **Co to:** Świadczeniodawca – może to być lekarz, szpital, klinika, laboratorium.
* **W datasetcie:** Obiekt predykcji – chcemy przewidzieć, czy dany provider jest oszukańczy (PotentialFraud).

**✅ Beneficjent (Beneficiary)**

* **Co to:** Osoba ubezpieczona w systemie Medicare, czyli pacjent.
* **W datasetcie:** Każdy BeneID reprezentuje jednego pacjenta. Jego dane zawarte są w pliku Beneficiarydata.

**✅ Roszczenie (Claim)**

* **Co to:** Wniosek o refundację kosztów leczenia złożony przez providera do Medicare.
* **W datasetcie:** Każde roszczenie zawiera informacje o leczeniu, procedurach, kosztach i przypisanym pacjencie.

**✅ Fraud (Oszustwo)**

* **Co to:** Świadome działania mające na celu nienależne uzyskanie pieniędzy z Medicare.
* **W datasetcie:** Kolumna PotentialFraud mówi, czy provider był oznaczony jako podejrzany.

**✅ Inpatient**

* **Co to:** Leczenie, podczas którego pacjent **zostaje w szpitalu** (np. na noc).
* **W datasetcie:** Plik Inpatientdata – zawiera roszczenia związane z hospitalizacjami (daty przyjęcia, wypisu itd.).

**✅ Outpatient**

* **Co to:** Leczenie **bez hospitalizacji** – np. wizyta u lekarza, badania ambulatoryjne.
* **W datasetcie:** Plik Outpatientdata – zawiera roszczenia związane z leczeniem bez pobytu w szpitalu.

**✅ Diagnosis Code (Kod diagnozy)**

* **Co to:** Standardowy kod (np. ICD-10) opisujący rozpoznanie choroby.
* **W datasetcie:** Występuje jako DiagnosisGroupCode, DiagnosisCode1, itd. – używany do uzasadnienia procedur.

**✅ Procedure Code (Kod procedury)**

* **Co to:** Kod opisujący wykonane świadczenie medyczne (np. operacja, badanie, zabieg).
* **W datasetcie:** Występuje jako ProcedureCode1, ProcedureCode2, itd. – powiązany z kosztami leczenia.

**✅ Refundacja (Reimbursement)**

* **Co to:** Kwota, jaką Medicare zwraca providerowi za wykonane świadczenie.
* **W datasetcie:** Kolumny takie jak InscClaimAmtReimbursed, IPAnnualReimbursementAmt itp.

**✅ Choroby przewlekłe (Chronic Conditions)**

* **Co to:** Długoterminowe schorzenia, np. cukrzyca, niewydolność serca, nowotwór.
* **W datasetcie:** Zawarte jako kolumny ChronicCond\_\* w pliku Beneficiarydata.

**✅ PotentialFraud**

* **Co to:** Etykieta klasyfikująca providera jako potencjalnie oszukańczego (Yes) lub nie (No).
* **W datasetcie:** Kluczowa zmienna celu (target) w pliku Train.csv.

**✅ Data urodzenia, śmierci (DOB, DOD)**

* **Co to:** Informacje o wieku i ewentualnym zgonie beneficjenta.
* **W datasetcie:** Używane do określenia wieku pacjenta i ew. ryzyka nadużycia (np. roszczenie po śmierci).